Niveau(x) d'enseignement: 4 Numéro EDUQ : 8270379
Centre d'intérêt: 74

Genesee, Fred.

Evaluation of the 1973-1974 grade VII French immersion class, February 1974 / Fred Genesee, T. Allister, S. Morin. -- [Montréal] : Commission des écoles protestantes du Grand Montréal, 1974.
54 f. ; 28 cm.

"A report submitted by the Curriculum Department to the Protestant School Board of Greater Montreal".

Cote Dewey: 440.7/G327e Cote LC: PC/2066

Descripteur(s): *Français (Langue) *Langue seconde *Classe d'immersion *Evaluation de l'apprentissage *Evaluation de programmes *Secondaire (Niveau)

L'étude vise à évaluer l'efficacité d'un programme d'immersion française mis sur pied en 1969-1970 et destiné à des élèves anglophones de septième année. Dans ce but, les performances scolaires (anglais, français et mathématiques) d'un échantillon d'élèves inscrits en 1973-1974 dans des classes d'immersion ont été comparées a celles de groupes de contrôle: élèves suivant l'enseignement régulier en anglais, élèves francophones. Le document rend compte d'une première série de résultats obtenue en février 1974.

Les 240 élèves de la région montréalaise qui se sont prêtés à l'enquête étaient répartis en quatre groupes de soixante personnes: immersion totale, immersion partielle, groupes témoins anglais et français. Les tests ont été passés en février 1974. Les instruments d'enquête suivants étaient utilisés: des extraits du "Stanford achievement test" et un test d'écriture en ce qui concerne la maîtrise de la langue anglaise; un test de lecture, un test d'écriture, le Test de rendement en français et le Test de rendement en mathématiques de la CECM, enfin un test de production et de compréhension orale pour ce qui est de l'apprentissage du français.

Comme par les années précédentes, les élèves des classes d'immersion ont effectué des progrès substantiels en langue française par rapport à leurs confrères des classes régulières, et celu sans que leur maîtrise de la langue anglaise en soit affectée. L'analyse phonétique des performances orales tend toutefois à indiquer que, sur ce point, il n'y a guère de différences entre les deux groupes. Une analyse grammaticale montre par ailleurs que les élèves des classes d'immersion ont davantage tendance à utiliser des structures syntaxiques complexes en français.

Type de document: r
Région(s) administrative(s): 063
Notice(s) connexe(s): 8270380, 8270377, 8270378, 8270380

Financement: Bénéficiaire - BEPGM
Organisme(s) source(s) - MEQM - DGEES
Programme(s) de subvention - PARC
Auteur de l'analyse: ar

E D U Q 8270379

EVALUATION OF THE 1973-1974

GRADE VII FRENCH IMMERSION CLASS

FEBRUARY 1974

A REPORT SUBMITTED BY THE CURRICULUM DEPARTMENT TO

THE PROTESTANT SCHOOL BOARD OF GREATER MONTREAL

Fred Genesee
T. Allister
S. Morin

September, 1974

E D U Q 8270379

EVALUATION OF THE 1973-1974

GRADE VII FRENCH IMMERSION CLASS

FEBRUARY 1974

A REPORT SUBMITTED BY THE CURRICULUM DEPARTMENT TO

THE PROTESTANT SCHOOL BOARD OF GREATER MONTREAL

Fred Genesee
T. Allister
S. Morin

September, 1974

CONTENTS

Page

Introduction 1

Summary of Previous Results 2

Current Assessment 3

Test Battery: 5

English Tests 5

French Tests 6

Results: 8

I.Q. 8

English Language Tests 10

French Language Tests 26

Discussion and Summary 45

INTRODUCTION

This report is the third in a series reporting the results of testing carried out to assess the effectiveness of French-Immersion instruction at the Grade VII Level.

History of Grade VII Immersion

In 1969-70 the Board inaugurated an experimental program of French-Immersion Instruction at the Grade VII level. This program was offered as an alternative to an early French Immersion program which starts in Kindergarten.

There are two parts to the "late immersion program". The Grade VII part involves a complete curriculum in French with the exception of daily English-language Arts classes. After completing Grade VII the students proceed through the higher grades with the option of taking a maximum of 40% of their curriculum in French.

In the first year of this program (1969-70) a select group of students participated. This group included only students whose French language skills were considered above average. Subsequently, participation in this program has been broadened to include any student requesting participation, and for whom there is no serious reason to suspect difficulty, such as learning difficulties or emotional problems. The 1970-71 Grade VII immersion group was evaluated in February and June, 1971, to assess the status of their English-language skills and, as well, to evaluate their progress in acquiring French. This group has been tested subsequently in Grades VIII, IX, and X, and will be tested again in the spring of this year, 1974, in their final year of high school.

An evaluation of a second Grade VII group was undertaken in November, 1971, and June, 1972. Test procedures for both projects were essentially the same.

SUMMARY OF PREVIOUS RESULTS

This summary will discuss briefly only the results from the Grade VII evaluations, and not the subsequent VIII, IX, and X evaluations.

Testing of the immersion children's English language skills indicated that they performed as well as, and in some instances better than, children attending the conventional English programs. Thus, it did not appear that one year's participation in an all-French program had had any detrimental effects on these children's native language skills.

The children were evaluated on French-language skills, using the following tests: listening comprehension, oral production, vocabulary, mathematics and reading. Their performance was compared to that of an English control group and a French control group. The results indicated that the immersion children performed better on all tests than did the English control group but did not perform as well as the French control group who were attending a French school. This pattern of results was expected.

With respect to the learning of content material, the immersion children performed at par with the English control children on the Arithmetic Concepts Subtest of the Standford Achievement Test, indicating that they were having no difficulty learning mathematical concepts in French.

A final area of assessment was related to attitudes towards oneself, English Canadians, French Canadians and European French persons. The results suggested that the Immersion children tended to see themselves "in a better light" than did the Control children. In terms of their attitudes towards French Canadians and European French the results revealed no differences of any significance between the groups of students tested. Thus, it appears that the French-immersion students had a relatively more favourable self-attitude but did not differ significantly from the English control students in terms of attitudes towards French Canadian or European French people.

SUMMARY OF PREVIOUS RESULTS

This summary will discuss briefly only the results from the Grade VII evaluations, and not the subsequent VIII, IX, and X evaluations.

Testing of the immersion children's English language skills indicated that they performed as well as, and in some instances better than, children attending the conventional English programs. Thus, it did not appear that one year's participation in an all-French program had had any detrimental effects on these children's native language skills.

The children were evaluated on French-language skills, using the following tests: listening comprehension, oral production, vocabulary, mathematics and reading. Their performance was compared to that of an English control group and a French control group. The results indicated that the immersion children performed better on all tests than did the English control group but did not perform as well as the French control group who were attending a French school. This pattern of results was expected.

With respect to the learning of content material, the immersion children performed at par with the English control children on the Arithmetic Concepts Subtest of the Standford Achievement Test, indicating that they were having no difficulty learning mathematical concepts in French.

A final area of assessment was related to attitudes towards oneself, English Canadians, French Canadians and European French persons. The results suggested that the Immersion children tended to see themselves "in a better light" than did the Control children. In terms of their attitudes towards French Canadians and European French the results revealed no differences of any significance between the groups of students tested. Thus, it appears that the French-immersion students had a relatively more favourable self-attitude but did not differ significantly from the English control students in terms of attitudes towards French Canadian or European French people.

CURRENT ASSESSMENT

In January, 1974, a third assessment of the Grade VII French Immersion program was set up. Four different types of schools were involved in the project: (1) French Immersion Centres: these are schools in which only the Grade VII French-Immersion Program is offered (Experimental I); (2) Winged Schools: these are schools which offer Grade VII French-Immersion as well as a regular English-language program (Experimental II): (3) English Control Schools: these are schools which offer only the English-language programs with French-as-a-second-language courses; and (4) French Control Schools: children from this sample were selected from the French high school in the P.S.B.G.M. (de Roberval) which offers a regular Grade VII to Grade XI curriculum in French. The French students took only some of the French language tests and none of the English language tests. The specific schools which were included in each category are enumerated in Table 1. These schools were selected to control for socioeconomic factors as much as possible. We were more successful in some cases than in others in this endeavour.

SAMPLE

In each group approximately half of the children were male and half were female. In the first three groups: (Experimental I, Experimental II and English Control) approximately one-third of the girls and one third of the boys had an average I.Q. score (95-115); one-third had below average I.Q. (80-94); and one-third had above average I.Q., (116+). The I.Q. measure that was available for the English students was the Lorge-Thorndike, which had been taken in 1970. Students who had a native French-speaking parent were excluded from the English samples. Children with any auditory or speech handicaps were excluded.

The verbal and non-verbal subtests of the Lavoie-Laurendeau Test Collectifs d'Intelligence Générals (1960) were administered to the French control sample in February, 1974. On the basis of their performance they were categorized as

TABLE 1

LIST OF SCHOOLS

EXPERIMENTAL I:

(Immersion Centre)

(1) Royal Vale
(2) Elmgrove

EXPERIMENTAL II:

(Winged Schools)

(1) Riverview
(2) Riverdale

ENGLISH CONTROL

(1) Hampstead
(2) Dalkeith
(3) Riverdale
(4) Riverview

FRENCH CONTROL

(1) de Roberval

Number of Students in Each Group:

Males:		Females:	
High I.Q.	... 10	High I.Q.	... 10
Average I.Q.	... 10	Average I.Q.	... 10
Low I.Q.	... 10	Low I.Q.	... 10

above average, average or below average in I.Q. Because these students had taken this test the previous year in Grade 6, their scores were somewhat inflated. Therefore, the criteria for inclusion into each of the I.Q. categories were increased by 5 points so that the average group comprised students with scores between 100-120; "above average" comprised students with a score of 121 or higher; and "below average" comprised students with a score of 99 or lower. The exact number of students in each sample is presented in Table 1.

TEST BATTERY

The two major areas of testing were French-language skills and English-language skills. No evaluation of the students' attitudes towards their own and other ethnolinguistic groups was taken at this time since a second testing in June of 1974 was to include such measures.

ENGLISH TESTS

1) Reading Test: The Reading Comprehension subtest of the California Achievement Tests (C.A.T.) was used here. This test consists of three prose passages graded in difficulty with a multiple-choice answer format. It is a timed test (40 min.). Raw scores from this test were converted into percentile rank. These percentiles are based on an American student population.

2) Writing Test: Each child was given a piece of paper on the top of which was typed the opening sentence of a composition which he was to complete. This test was untimed, and generally required from 30 to 40 minutes to complete. There were two different topics for the composition. Half of the students wrote on topic 1 and the other half wrote on topic 2. Both topics were translated into French. The topic which was not given to the child in English was later given to him in French.

3) Vocabulary: The Vocabulary subtest of the C.A.T. was used. This is a group test and is timed (10 minutes). It consists of 40 multiple-choice items. The child has to select one item out of a list of four which is synonomous with a reference word appearing at the head of the list. Again, raw scores were converted to standard percentiles based on an American population.

4) Spelling: Items from the Spelling subtest of the C.A.T. were presented orally to the students in groups. There were 29 items on this test. Scores were number of correct answers.

FRENCH TESTS

1) Reading Test: A passage entitled "J'ai parlé, Noa" from the children's magazine "Video Presse" (Vol. 1, No. 3) was presented to each child. Sixteen multiple-choice questions based on an edited version of the text were prepared. For each question the student had to select one of three alternatives which best matched the passage. The student was allowed to refer back to the passage while answering the questions. There was no time limit.

2) Writing: The same format and procedure as had been used in the English writing test was used here. As mentioned previously, there were two topics; the student was given one of the topics in English and the other in French. Approximately half of the students received topic 1 in French while the other half received topic 2 in French.

3) Test de Rendement en Français: This is a standardized test designed and published by the "Commission des Ecoles Catholiques de Montréal" (C.E.C.M.). The Grade V level was administered instead of Grade VII level since pretesting had shown that the Grade VII level was much too difficult. This test has 40 items which measure the following skills: spelling, word relevance, synonyms, antonyms, the function of words in sentences, gender and verb tenses. Raw scores were converted to stanine equivalents which are determined in November of each school year in the schools of the C.E.C.M.

4) Test de Rendement en Mathématiques: The Grade VII level of this test was administered. Again, this test is designed and standardized in schools of the C.E.C.M. Not all items are relevant for the students in the P.S.B.G.M. because of differences in curriculum. Thus, stanine derivatives of these scores are an underestimate of these students' mathematical achievement. Nevertheless, comparison

of the Experimental groups with the English control group using the raw scores gives an indication of these groups relative level of achievement in mathematics.

5) French oral production and oral comprehension: Samples of the children's oral French were collected by having them describe a story presented in the form of a cartoon sequence. After telling the story the child was then asked a number of informal questions related to the cartoon story and a number of pre-determined questions unrelated to the story. This portion of the interview was used to assess both production skills and listening skills. Some of these questions were designed to assess comprehension of certain structures; for example, subordination, logical propositions and simple mathematical calculations.

RESULTS

I.Q.:

The results from the French control group were not included in this analysis for two reasons. First of all, their results were not directly comparable to the results of the other three groups because they had taken a different intelligence test. Secondly, the results of the French control group seemed inflated possibly because they had taken the test in Grade 6 and again in the current year, Grade 7. To compensate for practice effects the criteria for the three French I.Q. groups were increased by 5 points as has been discussed already. Thus, there was a built-in difference between this group and the other three.

A three-way analysis of variance (least squares approximation for unequal n) was performed on the I.Q. scores. The independent variables were School Group (Experimental I, Experimental II and English Control), Sex (Male, Female) and I.Q. (average, below average, above average). See Table 2A for summary.

There was no significant difference between the Experimental I, Experimental II and English control group ($F = .99$; $df = 2, 161$; $p = .37$). The average I.Q. score for each group is presented in Table 2B. (see column sub-totals).

There was a significant difference between the three I.Q. groups ($F = 603.21$; $df = 2, 161$; $p = <.001$). This result simply confirms that the three groups we had selected so as to vary in terms of I.Q. did in fact differ in I.Q. The average I.Q. score for each I.Q. group is presented in Table 2B. (see row sub-totals).

There was no significant interaction between School Group and I.Q. group ($F = .09$; $df = 1, 161$; $p = .76$). The average I.Q. score for each of these sub-groups is presented in Table 2B. In other words, the I.Q.

TABLE 2A

THREE-WAY ANALYSIS OF VARIANCE FOR IQ SCORES

Source of Variation	df	MS	F Ratio	Sign.Level
Main Effects:				
Student Group	2	32.40	0.99	0.37
IQ Level	2	19798.1	603.21	0.00*
Sex of Student	1	3.04	0.92	0.76
Interaction Effects:				
Group x IQ	4	31.26	0.95	0.44
Group x Sex	2	13.10	0.40	0.67
IQ x Sex	2	21.15	0.64	0.53
Group x IQ x Sex	4	30.53	0.93	0.45
Experimental Error:	161	32.82		

TABLE 2B

SUMMARY OF MEAN I.Q. SCORES

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	TOTAL	FRENCH CONTROL
High I.Q.	124.25	122.70	121.3	122.75	129.5
Average I.Q.	105.59	103.02	104.87	104.50	112.0
Low I.Q.	86.10	87.26	85.44	86.27	87.5
TOTAL	105.32	104.33	103.87		

* $p < .001$

composition of each of the three school groups was comparable. The scores of the French control group have been included here for comparison. Remember that the range for each I.Q. category in this group was 5 points higher than in the other three groups. Thus, the range for the average I.Q. group was 100-120 (instead of 95-115), the range for the below average group was 99-(instead of 94-) and the range for the above average group was 120+ (instead of 116+).

English-Language Tests

Three separate three-way analyses of variance were performed on the scores from the spelling test, the vocabulary test and the reading test. The scores for the spelling test were in raw form, whereas the scores for the reading and vocabulary tests were in percentile form. The independent variables were School Group [Experimental I (immersion centres), Experimental II (winged schools), English Control Group], I.Q. (below average (94-), average (95-115) and above average (115+)) and Sex of student (male, female).

Spelling

A summary of the analysis of variance for the results from the Spelling test is presented in Table 3A. There were no significant differences between the three school groups: ($F = .65$; $df = 2,158$; $p = .52$). The average scores per group are presented in Table 3B (column sub-totals). The maximun possible score was 29.

There was a significant difference among the I.Q. groups ($F = 5.87$; $df = 2,158$; $p < .01$). The average scores per I.Q. group are presented in Table 3B (row sub-totals). The high I.Q. group had the highest score *($\bar{X} = 22.43$), followed by the average I.Q. group ($\bar{X} = 18.99$) and finally the low I.Q. group ($\bar{X} = 17.88$). This finding was not unexpected. The high I.Q.

*$\bar{X}$ = average

TABLE 3A

THREE-WAY ANALYSIS OF VARIANCE FOR SPELLING

Source of Variation	df	MS	F Ratio	Sign. Level
Main Effects:				
Student Group	2	36.52	0.65	0.52
I.Q. Level	2	328.92	5.87	0.004
Sex of Student	1	235.81	4.21	0.04
Interaction Effects:				
Group x I.Q.	4	19.14	0.34	0.85
Group x Sex	2	33.51	0.60	0.55
I.Q. x Sex	2	138.43	2.47	0.09
Group x I.Q. x Sex	4	122.21	2.18	0.07
Experimental Error:	158	56.01		

TABLE 3B

SUMMARY OF MEAN SPELLING SCORES

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	TOTAL
High I.Q.	23.15	21.90	22.25	22.43
Average I.Q.	19.33	19.39	18.24	18.98
Low I.Q.	19.44	16.02	18.19	17.88
TOTAL	20.64	19.10	19.56	

group scored significantly better than both the low I.Q. group (q = 4.74; df = 3, 158; p = .01)* and the average I.Q. group (q = 3.58; df = 2, 158); $p < .05$). There was no significant difference between the average and low I.Q. groups (q = 1.16; df = 2, 158; $p > .05$).

There was also a significant sex difference (F = 4.21; df = 1, 158; p = .04). The female students scored higher ($\bar{X}$ = 20.93) than did the male students ($\bar{X}$ = 18.61).

There was no significant School Group X I.Q. Group interaction (F = .34; df = 4, 158; p = .85). The means for each School Group at each I.Q. level are presented in Table 3B. A significant interaction means that performance depends on both variables, school group and I.Q. level in this case. An insignificant interaction indicates that the differences between the I.Q. groups is the same for each of the School Groups.

In summary, the differences found between the I.Q. groups, ignoring school program, are to be expected since tests of verbal skills generally correlate highly with tests of intelligence. The lack of any differences between the school groups suggests that no detrimental effects have resulted in English-spelling skills after five months' exposure to the French-immersion program.

Vocabulary

A summary of the analysis of variance for the vocabulary results is presented in Table 4A. There were no significant differences between the school groups (F = 2.44; df = 2, 161; p = .09). The average percentile scores per School Group are presented in Table 4B. (column sub-totals).

---

* Newman-Keuls multiple comparisons test Winer, 1971; p. 215 (See Footnote 1)

TABLE 4A

THREE-WAY ANALYSIS OF VARIANCE FOR VOCABULARY (CAT)

Source of Variation	df	MS	F Ratio	Sign. Level
Main Effects:				
Student Group	2	1037.52	2.44	0.09
I.Q. Level	2	12725.7	29.96	0.00*
Sex of Student	1	1298.60	3.06	0.08
Interaction Effects:				
Group x I.Q.	4	256.41	0.60	0.66
Group x Sex	2	345.12	0.81	0.45
I.Q. x Sex	2	762.48	1.80	0.17
Group x I.Q. x Sex	4	447.43	1.05	0.38
Experimental Error:	161	424.70		

TABLE 4B

SUMMARY OF MEAN VOCABULARY SCORES (PERCENTILE)

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	TOTAL
High I.Q.	85.50	83.75	76.50	81.92
Average I.Q.	73.23	71.60	61.58	68.80
Low I.Q.	56.80	48.92	54.44	52.72
TOTAL	71.84	68.09	63.50	

* $p < .001$

Again, there was a significant difference among the I.Q. groups ($F = 29.96$; $df = 2, 161$; $p < .001$). As expected, the high I.Q. group scored highest ($\bar{X} = 81.92$); the average I.Q. group was next ($\bar{X} = 68.80$); and the low I.Q. group scored lowest ($\bar{X} = 52.72$). All three groups scored significantly differently from one another (at the .01 level) by the Newman-Keuls procedure (Winer, 1971; p. 215).

There was no significant sex difference. Nor was there a significant School Group by I.Q. Group interaction ($F = .60$; $df = 4, 161$; $p = .66$). This latter finding indicates, as before, that the differences between the I.Q. groups are the same for each school group. See Table 4B for the average vocabulary score for each I.Q. group within each school group.

Thus, there is no evidence that the experimental students are suffering in terms of vocabulary skills after five months in the program. With a combined average percentile score of 70% the two experimental groups are scoring higher than 70% of the American sample who have taken this test.

Reading

A summary of the analysis of variance for the reading scores is presented in Table 5A. There was a significant difference among the School Groups ($F = 9.74$; $df = 2, 161$; $p < .001$). The Experimental I group scored highest ($\bar{X} = 66.20$) ; The Experimental II Group scored next highest ($\bar{X} = 50.29$); and the English Control Group scored lowest ($\bar{X} = 49.33$). There is virtually no difference in mean scores between the Experimental II group and the Control Group.

These findings are somewhat puzzling since they have not been found before. The possibility of these differences being due to differences in intelligence must be ruled out since the three groups were equated on I.Q. Another possible explanation is motivation, both while taking the reading

TABLE 5A

THREE-WAY ANALYSIS OF VARIANCE FOR READING (CAT)

Source of Variation	df	MS	F Ratio	Sign. Level
Main Effects:				
Student Group	2	5390.75	9.74	0.00*
IQ Level	2	14871.5	26.88	0.00*
Sex of Student	1	1158.64	2.09	0.15
Interaction Effects:				
Group x IQ	4	524.51	0.95	0.44
Group x Sex	2	403.20	0.73	0.48
IQ x Sex	2	211.28	0.38	0.68
Group x IQ x Sex	4	110.22	0.20	0.93
Experimental Error:	161	553.27		

TABLE 5B

SUMMARY OF MEAN READING SCORES (PERCENTILE)

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	TOTAL
High IQ	80.15	70.90	59.10	70.05
Average IQ	67.53	48.12	56.07	57.24
Low IQ	51.15	31.87	32.81	30.61
TOTAL	66.28	50.30	49.33	

* $p < .001$

test and in school generally. The testers reported that the control students were not as enthusiastic or favorable towards the testing as were the experimental students. While the difference in motivation was not extreme, it was nevertheless noticeable. Secondly, it is possible that because of their somewhat "special status" in school, the immersion children are more motivated and enthusiastic generally toward academic events. This possibility bears further investigation. The final possible explanation, although the least likely, is that, in fact, taking the Grade 7 curriculum in French improves English reading skills.

These results also give some indication of these students' standing relative to a larger American population. Both the Control Group and the Experimental II Group scored at approximately the 50th percentile while the Experimental I Group scored at the 66th percentile. Combining all groups, the average percentile standing is 55. This means that, in general, this sample of students is reading better than 55% of a "standard American sample".

Again, there was a significant difference among the I.Q. groups ($F = 26.88$; $df = 2, 161$; $p < .001$). The average scores per I.Q. group are presented in Table 5B. The high I.Q. group ($\bar{X} = 70.05$) scored significantly better than both the average I.Q. group with an average score of 57.24 ($q = 4.21$; $df = 2, 161$; $p < .01$)* and the low I.Q. group with an average score of 39.61 ($q = 10.34$; $df = 3, 161$; $p < .01$). As well, the average I.Q. group scored significantly higher than the low I.Q. group ($q = 6.12$; $df = 2, 161$; $p < .01$).

There were no other significant effects or interactions. See Table 5B for a summary of the scores by school and I.Q. group.

---

* Newman-Keuls multiple-comparison tests, Winer, 1972, p.215.

In summary, there is no evidence that after approximately five months in the Grade VII French-immersion program the students have suffered any detrimental effects in their spelling, vocabulary or reading skills in English. In fact, there was some indication that students in the French-immersion centres were reading significantly better than either the Control students or the students in the French-immersion winged-schools. The exact reason for this latter result is not entirely clear at this time.

English Compositions

Each composition was scored by two grade 7 teachers; one taught English to students in French-immersion and the other taught in an all-English school. All compositions were mixed together and were scored blind on nine different measures; namely, spelling, punctuation, length, sentence accuracy, sentence complexity & variety, organization, vocabulary, originality and overall. See Appendix A for the rating scales and criteria that were used.

Six separate four-way analyses of variance were performed on the ratings for spelling, sentence accuracy, sentence complexity & variety, organization, originality and overall. The main variables were School group, IQ group, Sex of student and Marker. Frequency counts were made of the ratings for punctuation, length and vocabulary.

Spelling: The average rating per school group, IQ group and sex is presented in Table 6A. The only significant effect was due to IQ ($F = 10.53$; $df = 2, 100$; $p = <.05$). As expected, the high IQ group had the highest mean rating ($\bar{X} = 2.9$). There were no significant differences among the School groups. ($F = 1.70$; $df = 2, 21$; $p > .05$). The combined group average was 3.3 which means that on the average the children were making 4 to 6 spelling errors.

Sentence Accuracy: The average ratings per School group, IQ group and Sex of student are presented in Table 6B. There were no significant

TABLE 6A

	EXPERIMENTAL I		EXPERIMENTAL II		ENGLISH CONTROL		
	Male	Female	Male	Female	Male	Female	TOTAL
HIGH I.Q.	3.6	3.5	3.7	3.6	3.1	3.6	3.67
AVERAGE I.Q.	2.7	3.3.	2.8	2.9	2.9	3.0	3.35
LOW I.Q.	2.8	2.7	2.4	2.9	2.3	2.4	2.91
TOTAL GROUP	3.26		3.49		3.17		

Summary of Mean Ratings on Spelling for each School, I.Q. and Sex Group

TABLE 6B

	EXPERIMENTAL I		EXPERIMENTAL II		ENGLISH CONTROL		
	Male	Female	Male	Female	Male	Female	TOTAL
HIGH I.Q.	3.6	3.5	3.7	3.6	3.1	3.6	3.48
AVERAGE I.Q.	2.7	3.3	2.8	2.9	2.9	3.0	2.94
LOW I.Q.	2.8	2.7	2.4	2.9	2.3	2.4	2.59
TOTAL GROUP	3.10		3.05		2.87		

Summary of Mean Ratings on Sentence Accuracy for each School, I.Q. and Sex Group

differences between the School groups ($F = 1.3$; $df = 2, 51$; $p > .05$). With a combined group average of 3.00 the students were rated as meeting "the objectives of grade seven level".

There was a significant effect due to IQ ($F = 18.73$; $df = 2, 66$; $p < .05$). The high IQ group had the highest mean rating ($\bar{X} = 3.48$), followed by the average IQ group ($\bar{X} = 2.94$) and finally the low IQ group ($\bar{X} = 2.58$). There was a significant difference in the ratings of each marker ($F = 6.65$; $df = 1, 57$; $p = .01$). Marker 1 ($\bar{X} = 2.9$) tended to give lower ratings than marker 2 ($\bar{X} = 3.1$). The difference, although statistically significant, is small indeed, and both markers' ratings place the students at a "satisfactory" level.

There was a Marker X Sex of Student interaction ($F = 12.77$; $df = 1, 157$; $p = <.001$). Marker 1 rated the boys ($\bar{X} = 2.76$) lower than the girls ($\bar{X} = 3.11$), while Marker 2 rated the girls ($\bar{X} = 3.06$) lower than the boys ($\bar{X} = 3.10$). Incidentally, Marker 1 was a male and Marker 2 was a female. This interaction recurs in other analyses.

Sentence Complexity & Variety: A summary of the average ratings per group is presented in Table 7A. There was a significant School group effect ($F = 4.9$; $df = 2, 81$; $p = <.05$). The Experimental I group ($\bar{X} = 3.03$) had a higher mean rating than either the Experimental II group ($\bar{X} = 2.98$) or the English Control group ($\bar{X} = 2.68$). Although the differences appear to be small they are reliable and consequently statistically significant. The average rating of the two experimental groups combined, places them in the category "satisfactory for grade seven level" while the average rating for the English Control students suggests that their writing "needs work" in terms of sentence complexity & variety.

There was also a significant main effect due to IQ ($F = 9.8$; $df = 2, 10$; $p < .05$). Again, as expected, the high IQ group had the highest rating ($\bar{X} = 3.28$), followed by the average IQ group ($\bar{X} = 2.83$) and finally the low IQ group ($\bar{X} = 2.59$).

TABLE 7A

	EXPERIMENTAL I		EXPERIMENTAL II		ENGLISH CONTROL		
	Male	Female	Male	Female	Male	Female	TOTAL
HIGH I.Q.	3.5	3.3	3.3	3.4	2.8	3.4	3.28
AVERAGE I.Q.	2.5	3.2	2.9	2.8	2.8	2.9	2.83
LOW I.Q.	2.9	2.9	2.6	2.9	2.1	2.2	2.59
TOTAL GROUP	3.03		2.98		2.68		

Summary of Mean Ratings on Sentence Complexity & Variety for each School, I.Q. and Sex Group

TABLE 7B

	EXPERIMENTAL I		EXPERIMENTAL II		ENGLISH CONTROL		
	Male	Female	Male	Female	Male	Female	TOTAL
HIGH I.Q.	3.3	3.3	3.2	3.3	2.7	3.2	3.15
AVERAGE I.Q.	2.4	3.2	2.6	2.7	2.7	2.6	2.69
LOW I.Q.	3.0	2.4	2.3	2.4	2.0	2.3	2.39
TOTAL GROUP	2.92		2.74		2.57		

Summary of Mean Ratings on Organization for each School, I.Q. and Sex Group

There was a significant Marker by Sex of student interaction ($F = 7.97$; $df = 1, 157$; $p = <.001$). Marker 1 gave higher ratings to the girls ($\bar{X} = 3.06$) than to the boys ($\bar{X} = 2.68$) while Marker 2 gave the same mean rating to both boys ($\bar{X} = 2.92$) and girls ($\bar{X} = 2.92$).

Organization: A summary of the mean ratings for organization is presented in Table 7B. There were no significant differences among the School groups. There was a significant effect due to IQ ($F = 23.84$; $df = 2, 57$; $p = <.05$). The high IQ group had the highest mean rating ($\bar{X} = 3.15$), followed by the average IQ group ($\bar{X} = 2.69$) and finally the low IQ group ($\bar{X} = 2.39$).

There was a significant interaction between Marker and Sex of Student ($F = 5.63$; $df = 1, 157$; $p = <.05$). Marker 1 gave higher ratings to the girls ($\bar{X} = 2.71$) than to the boys ($\bar{X} = 2.46$) while Marker 2 gave the same mean rating to the boys ($\bar{X} = 2.91$) and the girls ($\bar{X} = 2.90$).

Originality: A summary of the mean ratings on originality for each group is presented in Table 8A. There were no significant differences between the School groups ($F = 2.94$; $df = 2, 8$; $p > .05$). The mean combined group rating of 2.03 suggests that these markers found the compositions, on the average, to be "ordinary".

There was a significant main effect due to IQ ($F = 12.8$; $df = 2, 49$; $p = <.05$). The high IQ group had the highest mean rating ($\bar{X} = 2.28$), followed by the average IQ group ($\bar{X} = 1.98$) and finally the low IQ group ($\bar{X} = 1.84$).

There was a significant difference in the ratings of the two Markers ($F = 4.5$; $df = 1, 157$; $p = <.05$). Marker 1 gave less favorable ratings ($X = 1.98$) than did Marker 2 ($\bar{X} = 2.09$). However, this difference is very small.

Again, there was a significant interaction between Marker and Sex of Student ($F = 5.41$; $df = 1, 157$; $p = <.05$). Marker 1 gave higher mean ratings to the girls ($\bar{X} = 2.02$) than to the boys ($\bar{X} = 1.93$) while Marker 2 gave higher mean ratings to the boys ($\bar{X} = 2.17$) than to the girls ($\bar{X} = 2.01$).

TABLE 8A

	EXPERIMENTAL I		EXPERIMENTAL II		ENGLISH CONTROL		
	Male	Female	Male	Female	Male	Female	TOTAL
HIGH I.Q.	2.5	2.4	2.4	2.3	2.0	2.3	2.28
AVERAGE I.Q.	1.9	2.3	2.0	2.0	1.9	1.8	1.98
LOW I.Q.	2.2	1.8	1.9	1.7	1.7	1.7	1.84
TOTAL GROUP	2.17		2.03		1.89		

Summary of Mean Ratings on Originality for each School, I.Q. and Sex Group

TABLE 8B

	EXPERIMENTAL I		EXPERIMENTAL II		ENGLISH CONTROL		
	Male	Female	Male	Female	Male	Female	TOTAL
HIGH I.Q.	3.5	3.3	3.4	3.4	2.8	3.3	3.24
AVERAGE I.Q.	2.3	3.0	2.9	2.9	2.7	2.7	2.86
LOW I.Q.	2.5	2.6	2.3	2.7	2.0	2.2	2.40
TOTAL GROUP	3.86		2.92		2.63		

Summary of the Mean Overall Ratings for each School, I.Q. and Sex Group

Overall: A summary of the main overall ratings per group is presented in Table 8B. The main School group difference was marginally significant ($F = 3.1$; $df = 2, 70$; $p = <.06$). The Experimental I ($\bar{X} = 2.86$) and the Experimental II ($\bar{X} = 2.92$) group tended to have higher mean overall ratings than the English control students ($\bar{X} = 2.63$). While these differences are small they are nevertheless reliable. The combined group average of 2.8 suggests that overall the compositions are "average" (or warrant a C grade).

There was a significant effect due to I.Q. ($F = 22.69$; $df = 2, 64$; $p = <.05$). The high I.Q. group had the highest mean overall rating ($\bar{X} = 3.24$), followed by the average I.Q. group ($\bar{X} = 2.76$) and finally the low I.Q. group ($\bar{X} = 2.4$).

There was a significant interaction between Marker and Sex of Student ($F = 7.97$; $df = 1, 137$; $p = <.05$). Marker 1 gave higher ratings to the girls ($\bar{X} = 2.93$) than to the boys ($\bar{X} = 2.61$) while Marker 2 did not differentiate much between the boys ($\bar{X} = 2.80$) and the girls ($\bar{X} = 2.86$).

Summary

On all six variables there were significant I.Q. effects with the high I.Q. group being rated highest and the low I.Q. group lowest. On five of the six variables there were significant Marker by Sex of Student interactions; these are difficult to explain and not really pertinent to the question of the effectiveness of French-immersion.

On two of the variables (Sentence Complexity and Accuracy and Overall) there were significant School group differences. In both cases the two experimental groups were rated more highly than the control group.

Thus, five months of French-immersion does not seem to have harmed the students' English writing skills. On the contrary, on the two variables where there were differences they were in favour of the French-immersion group. Whether these differences are in fact due to the fact that French is the medium of instruction

or to other related factors, such as differential motivation, is difficult to say. These other factors are nevertheless often concomitants of the French-immersion experience and cannot be discounted as irrelevant.

Punctuation, Length and Vocabulary:

Because there were only two possible ratings for the Punctuation and Length scores (Appropriate or Inappropriate) and only three for the Vocabulary score (Superior, Average or Below Average) the results have been summarized in percentage frequency tables (see Table 9). The percentages have been calculated ignoring I.Q., sex and marker distinctions.

It is apparent from these results that the two immersion groups have scored as well as or better than the Controls. On Punctuation the Experimental I group and the English Control group have comparable percentages in both the "Appropriate" and "Inappropriate" categories. The Experimental II group seems to be doing slightly better than the other two groups in that they have a higher percentage of scores in the "Appropriate" category.

On Length the two Experimental groups have been scored similarly, with two-thirds of the ratings being "Appropriate", while the English Control group had a higher percentage of scores in the "Inappropriate" category than in the "Appropriate" category.

On Vocabulary both Experimental groups have a higher percentage of scores in the "Superior" category and fewer in the "Below Average" category than did the English Controls.

Again, there is no reason for concern - the French-immersion group are writing as well as the English Control group. In fact, there is some suggestion that they are doing slightly better.

TABLE 9

RESULTS OF ANALYSIS OF ENGLISH COMPOSITIONS ON PUNCTUATION, LENGTH AND VOCABULARY IN PERCENTAGE FREQUENCY

	PUNCTUATION		LENGTH		VOCABULARY		
	Appropriate	Inappropriate	Appropriate	Inappropriate	Superior	Average	Below Average
Experimental I	63.5%	36.5%	65.5%	34.5%	12%	77.5%	10.5%
Experimental II	75.5%	24.5%	65.5%	34.5%	9.5%	80.5%	10%
English Control	62.5%	37.5%	43%	57%	7.5%	71.5%	21%

French-Language Tests

Five separate three-way analyses of variance were performed on the following scores: reading (raw form), Test de Rendement en Français (raw form and stanine), Test de Rendement en Mathématiques (raw form and stanine). Again, the independent variables were School Group (Experimental I, Experimental II, English Control and French Control), I.Q. and Sex of Student.

Reading: A summary of the analysis of variance for the French reading scores is presented in Table 10A. There was a significant main effect due to School Group ($F = 21.41$; $df = 3,209$; $p = <.001$). The average score per group is presented in Table 10B. The maximum possible score on this test is 16. A series of Newman-Keuls tests (Winer, 1971; p. 215) revealed that the French Control group ($\bar{X} = 12.10$), scored significantly better than the Experimental I ($\bar{X} = 7.38$), the Experimental II ($\bar{X} = 7.38$), and the English Control group ($\bar{X} = 5.97$). The two experimental groups did not differ significantly from one another, and only the Experimental I Group scored significantly higher than the English Control group. Although the Experimental II group had an average total score greater than that of the English Control group this difference did not reach statistical significance at the .05 level. ($q = 2.47$; $df = 2, 209$; $p = .08$). Chance performance on this test is a score of 5.33; there are 16 questions with 3 alternatives for each one. Thus, the English Control group with an average score of 5.97 is scoring at chance level while the two experimental groups are scoring above chance.

There was a significant difference in performance due to I.Q. ($F = 5.54$; $df = 2, 209$; $p = .005$). The average scores per I.Q. group are presented in Table 10B. The high I.Q. group ($\bar{X} = 9.66$) performed significantly better than the low I.Q. group ($\bar{X} = 7.36$) but did not perform significantly better than the average I.Q. group ($\bar{X} = 8.79$) (Newman-Keuls multiple-comparison tests; Winer, 1971; p. 215). The average I.Q. group scored significantly higher than the low I.Q. group.

TABLE 10A

THREE-WAY ANALYSIS OF VARIANCE FOR FRENCH READING TEST

Source of Variation	df	MS	F Ratio	Sign. Level
Main Effects:				
Student Group	3	400.41	21.41	0.000
IQ Level	2	103.54	5.54	0.005
Sex of Student	1	15.75	0.84	0.36
Interaction Effects:				
Group x IQ	6	39.00	2.09	0.056
Group x Sex	3	11.06	0.59	.62
IQ x Sex	2	22.02	1.18	0.31
Group x IQ x Sex	6	9.65	0.52	0.80
Experimental Error:	209	18.70		

TABLE 10B

Summary of Mean Scores on the French Reading Test

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	FRENCH CONTROL	TOTAL
High IQ	9.25	8.65	6.40	14.33	9.66
Average IQ	10.64	6.46	5.39	12.65	8.79
Low IQ	7.00	7.02	6.11	9.31	7.36
TOTAL	8.96	7.38	5.97	12.09	

The School Group by I.Q. Group interaction was significant ($F = 2.09$; $df = 6, 209$; $p = .056$). The average score for each I.Q. group within each school group is presented in Table 6B. In general, a significant interaction indicates that performance on the test depends on both variables, I.Q. and school group in this instance. One possible interpretation of this interaction is as follows: the reading scores of the francophone students in the French Control group correlate positively with I.Q.; in other words, the higher one's I.Q., the higher one's reading score. On the other hand, for the anglophone students from the other three groups (Experimental I, Experimental II, and English Control) there is no similar positive correlation between reading performance and I.Q.; in other words, results on the French reading test are not necessarily predictable from I.Q. score. For example, in the Experimental I group the highest score was made by the average I.Q. group, while in the Experimental II group the highest score was made by the high I.Q. group.

In summary, it is evident that after five months in the French-immersion program the students were reading French better than their peers in the English program but not as well as native French-speaking Grade VII students. This was expected. There was some suggestion that reading performance in French was not predictable from I.Q. score; this finding warrants further investigation.

Test de Rendement en Français: A summary of the analysis of variance of the raw scores from the Test de Rendement en Français (Grade V level) is presented in Table 11A. There was a significant difference in performance among the four groups ($F = 139.60$; $df = 3, 203$; $p < .001$). The average raw score per School Group is presented in Table 11B. The maximum possible score on this test is 35. Newman-Keuls multiple-comparison tests (Winer, 1971; p. 215) revealed that every group differed significantly from every other group at the .01 level. Thus, the French control group performed significantly better than every other

TABLE 11A

THREE-WAY ANALYSIS OF VARIANCE FOR Test de Rendement en Français
(Raw Scores)

Source of Variation	df	MS	F Ratio	Sign. Level
Main Effects:				
Student Group	3	2843.29	139.60	0.000
IQ Level	2	591.38	29.04	0.000
Sex of Student	1	96.29	4.73	0.03
Interaction Effects:				
Group x IQ	6	63.17	3.10	0.006
Group x IQ	3	11.86	0.58	0.63
IQ x Sex	2	21.94	1.08	0.34
Group x IQ x Sex	6	4.64	0.23	0.97
Experimental Error:	209	20.37		

TABLE 11B

Summary of the Mean Scores on the Test de Rendement en Français

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	FRENCH CONTROL	TOTAL
High I.Q.	16.90	16.85	11.20	30.54	18.87
Average I.Q.	15.63	13.28	9.16	27.90	16.49
Low I.Q.	13.44	10.16	8.69	20.64	13.23
TOTAL	15.32	13.43	9.68	26.36	

group, as expected. The two experimental groups performed significantly better than the English Control Group, and the two experimental groups differed significantly from one another with the Experimental I group scoring higher than the Experimental II group.

Converting the raw scores to stanines gives an indication of the performance of the various groups relative to the Grade V francophone students in schools of the C.E.C.M. as of November,1973. The average stanine score per school group is presented in Table 12A (see column sub-totals). The statistical analysis of these scores is the same as that of the raw scores and, thus, will not be discussed. An interpretation of stanine scores is presented as Appendix B. With an average stanine of 6.7 the French Control group scored at a level comparable to the top stanine of the average portion of the Grade V francophone class. With a combined average stanine of 3.35 the two French-immersion groups scored at a level comparable to the top stanine of the below average portion of the Grade V francophone class. Finally, with an average stanine of 2.08 the English control students scored at a level comparable to the bottom stanine of the below average portion of the Grade V francophone class.

Returning to the analysis of the raw scores, there was a significant difference among the I.Q. groups ($F = 29.04$; $df = 2, 203$; $p < .001$). The average score per I.Q. group is presented in Table 11B. As expected, the high I.Q. group had the highest score ($\bar{X} = 18.87$), the low I.Q. group had the lowest score ($\bar{X} = 13.23$) and the average I.Q. group was intermediate ($\bar{X} = 16.49$). Newman-Keuls multiple-comparison tests revealed that all groups differ significantly from all other groups.

There was also a significant main effect due to Sex of student ($F = 4.73$; $df = 1, 203$; $p = .03$). The girls scored higher ($\bar{X} = 16.86$) than did the boys ($\bar{X} = 15.54$).

TABLE 12A

THREE-WAY ANALYSIS OF VARIANCE FOR the Test de Rendement en Français (Stanines)

Source of Variation	df	MS	F Ratio	Sign. Level
Main Effects:				
Student Group	3	218.59	130.31	.000
IQ Level	2	42.35	25.15	.000
Sex of Student	1	6.21	3.7	.06
Interaction Effects:				
Group x IQ	6	5.34	3.18	.005
Group x Sex	3	2.97	0.38	.63
IQ x Sex	2	2.21	1.32	.27
Group x IQ x Sex	6	0.23	0.14	.99
Experimental Error:	209	1.68		

TABLE 12B

Summary of the Mean Stanine Scores on the Test de Rendement en Français

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	FRENCH CONTROL	TOTAL
High I.Q.	3.9	3.9	2.5	7.9	4.58
Average I.Q.	3.7	3.3	1.9	7.1	4.00
Low I.Q.	3.1	2.3	1.8	5.1	3.07
TOTAL	3.57	3.17	2.09	6.71	

There was a significant interaction between School Group and I.Q. Group ($F = 3.10$; $df = 6, 203$; $p = .006$). The average scores for each of these samples is presented in Table 11B. This interaction is not easily interpreted.

In summary, the French-immersion students performed better than the English control group but not as well as the French control group on this test. Relative to the norms of "La Commission des Ecoles Catholiques" the French-immersion groups performed at a level comparable to the top stanine (i.e. 3) portion of the below average francophone Grade V class. The English control group performed at a level comparable to the bottom stanine (i.e. 2) of the below average francophone Grade V class.

Test de Rendement en Mathématiques: A summary of the analysis of variance of the raw scores from the Test de Rendement en Mathématiques (level 7) is presented in Table 13A. There was a significant difference among the school groups ($F = 24.33$; $df = 3, 209$; $p < .001$). The average raw scores are presented in Table 13B. The most noteworthy finding is that the Experimental I group ($\bar{X} = 19.86$) had the highest score on this test, higher even than the French control group ($\bar{X} = 17.91$). In fact, this difference was statistically significant (Newman-Keuls test, Winer, 1971; p.215). The Experimental I Group also scored significantly higher than the English Control group ($\bar{X} = 13.26$) which had the lowest score of all four groups. The Experimental II Group and the French control group both scored significantly higher than the English control group.

The superiority of the Experimental group over the French control group on the French mathematics test cannot be attributed to differences in I.Q. since I.Q. was controlled, unless the adjustment that was made in the criteria for the French group was not appropriate. Differences in test conditions may have worked in favor of the Experimental group.

TABLE 13A

THREE-WAY ANALYSIS OF VARIANCE FOR the Test de Rendement en Mathématiques (Raw Scores)

Source of Variation	df	MS	F Ratio	Sign. Level
Main Effects:				
Student Group	2	474.44	24.33	.000
IQ Level	2	884.40	45.36	.000
Sex of Student	1	103.27	5.30	.02
Interaction Effects:				
Group x IQ	4	24.00	1.23	.29
Group x Sex	2	18 99	.97	.41
IQ x Sex	2	20.06	1.03	.36
Group x IQ x Sex	4	27.40	1.41	.21
Experimental Error:	209	19.50		

TABLE 13B

Summary of the Mean Scores on the Test de Rendement en Mathématiques

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	FRENCH CONTROL	TOTAL
High I.Q.	22.65	23.35	15.95	20.59	20.63
Average I.Q.	19.42	18.42	14.07	18.65	17.64
Low I.Q.	17.50	13.73	9.77	14.50	13.88
TOTAL	19.86	18.50	13.26	17.91	

In terms of stanine scores (see Table 14B) the Experimental I and II groups had a mean stanine of 5.4 and 5.0, respectively, which places them at an average level of performance compared to the Grade 7 francophone students in the C.E.C.M. The French Group had a stanine of 4.8 which is also average. The English Control group scored in the top stanine of the below average group.

Returning to the analysis of the raw scores, there was a significant difference among the I.Q. groups ($F = 45.35$; $df = 2, 209$; $p < .001$). The high I.Q. group had the highest average score ($\bar{X} = 20.63$), followed by the average I.Q. group ($\bar{X} = 17.64$) and finally the low I.Q. group ($\bar{X} = 13.88$).

There was a significant difference between the male and female students ($F = 5.3$; $df = 1, 209$; $p = .02$) with the boys ($\bar{X} = 18.05$) scoring higher than the girls ($\bar{X} = 16.71$).

Thus, it is evident that receiving mathematics instruction in French has not hampered the French-immersion students in acquiring mathematical concepts: their performance on the standardized test of the C.E.C.M. places them above the francophone sample that was used in this assessment as a comparison. The experimental students scored decidedly better than their English counterparts from the English program. The experimental students scored at average level for the Grade VII francophone population while the English Control children scored below average.

French Compositions:

Each composition was scored by two grade seven teachers; one taught in a French-immersion program and the other taught in an all-French elementary school. All compositions were mixed together and were scored blind on the following five dimensions: sentence accuracy, vocabulary, spelling 1 (orthographe d'usage), spelling 2 (orthographe grammaticale), organization and overall. See Appendix C for a copy of the rating scales and criteria that were used.

TABLE 14A

THREE-WAY ANALYSIS OF VARIANCE FOR the Test de Rendement en Mathématiques (Stanines)

Source of Variation	df	MS	F Ratio	Sign. Level
Main Effects:				
Student Group	2	50.50	24.18***	.000
IQ Level	2	94.33	45.17***	.000
Sex of Student	1	14.48	6.93**	.009
Interaction Effects:				
Group x IQ	4	2.78	1.33	.24
Group x Sex	2	1.31	.63	.60
IQ x Sex	2	2.43	1.64	.31
Group x IQ x Sex	4	2.18	1.04	.40
Experimental Error:	203	2.09		

*P < 0.05 **P < 0.01 ***P < 0.001

TABLE 14B

Summary of the Mean Stanine Scores on the Test de Rendement en Mathématiques

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	FRENCH CONTROL	TOTAL
High I.Q.	6.3	6.7	4.3	5.7	5.7
Average I.Q.	5.3	4.9	3.6	5.0	4.7
Low I.Q.	4.7	3.5	2.0	3.8	3.5
TOTAL	5.4	5.0	3.3	4.8	

Six separate four-way analyses of variance were performed on the ratings. The main variables were School group (Experimental I, Experimental II, English Control and French Control), IQ group, Sex of student and Marker. Frequency counts were made of the grammatical faults listed in item 1B, Appendix C.

Vocabulary: The average rating per School group and IQ group is presented in Table 15A. There was a significant effect due to School Group ($F = 17.33$; $df = 3, 4$; $p < .01$). The French Control group was rated highest ($\bar{X} = 4.02$), followed by the two experimental groups (Experimental I = 2.72; Experimental II = 2.53) and finally the English Control group ($\bar{X} = 1.66$).

There was a significant effect due to IQ ($F = 18.24$; $df = 2, 108$; $p < .01$). The high IQ group was rated highest ($\bar{X} = 3.05$), followed by the average IQ group ($\bar{X} = 2.63$) and finally the low IQ group ($\bar{X} = 2.51$).

There was a significant difference between Marker 1 and Marker 2 ($F = 39.01$; $df = 1, 206$; $p < .001$). Marker 2 ($\bar{X} = 2.87$) tended to give higher ratings than did Marker 1 ($\bar{X} = 2.59$).

There was a significant Marker by Sex of student interaction ($F = 4.98$; $df = 1, 206$; $p < .05$). Marker 1 gave higher ratings to the girls ($\bar{X} = 2.78$) than to the boys ($\bar{X} = 2.40$) while Marker 2 did not differentiate between them ($\bar{X}_{(boys)} = 2.78$; $\bar{X}_{(girls)} = 2.96$).

Sentence Accuracy: A summary of the mean ratings for each School and IQ group is presented in Table 15B. There was a significant difference between the School groups ($F = 101.42$; $df = 3, 62$; $p < .001$). The French group ($\bar{X} = 3.52$) had the highest mean rating, then the two experimental groups ($\bar{X}_I = 2.59$; $\bar{X}_{II} = 2.47$) and finally the English Control group ($\bar{X} = 1.70$).

There was a significant difference between the IQ groups ($F = 9.29$; $df = 2, 8$; $p = < .01$). The high IQ group had the highest mean rating ($X = 2.90$), followed by the average IQ group ($\bar{X} = 2.51$) and finally the low IQ group ($\bar{X} = 2.29$).

TABLE 15A

SUMMARY OF RATINGS ON VOCABULARY FOR FRENCH COMPOSITIONS

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	FRENCH CONTROL	TOTAL
High I.Q.	3.00	2.85	1.88	4.48	3.05
Average I.Q.	2.65	2.40	1.58	3.88	2.63
Low I.Q.	2.50	2.33	1.51	3.71	2.51
TOTAL	2.71	2.53	1.66	4.02	

TABLE 15B

SUMMARY OF RATINGS ON SENTENCE ACCURACY FOR FRENCH COMPOSITIONS

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	FRENCH CONTROL	TOTAL
High I.Q.	2.75	2.93	1.88	4.06	2.90
Average I.Q.	2.64	2.33	1.67	3.41	2.51
Low I.Q.	2.39	2.14	1.56	3.08	2.29
TOTAL	2.60	2.47	1.70	3.52	

TABLE 15C

SUMMARY OF RATINGS ON SPELLING I FOR FRENCH COMPOSITIONS

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	FRENCH CONTROL	TOTAL
High I.Q.	2.38	2.50	1.73	2.86	2.37
Average I.Q.	2.17	2.12	1.58	2.62	2.12
Low I.Q.	2.16	2.08	1.50	2.45	2.05
TOTAL	2.23	2.23	1.60	2.64	

There was a significant sex difference ($F = 22.89$; $df = 1, 113$; $p < .001$). The girls ($\bar{X} = 2.72$) had higher ratings than did the boys ($\bar{X} = 2.42$).

There was no significant difference between the Markers' ratings.

Spelling 1 (Orthographe d'usage): Spelling 1 type errors comprise errors in the spelling of invariant words, such as nouns, prepositions and conjunctions. Spelling 2 type errors comprise errors in the spelling of words that must agree grammatically with other words or structures in the sentence, such as adjectives, past participles, verbs, etc. For example, la mayson is a type 1 error, whereas la maison blanc is a type 2 error since the adjective should be blanche, the feminine of blanc; this is a grammatical error.

A summary of the type 1 errors for each School and IQ group is presented in Table 15C. There was a significant difference between the School groups ($F = 61.36$; $df = 3, 23$; $p = < .01$). The French group made fewest errors ($\bar{X} = 2.64$), followed by the two experimental groups ($\bar{X}I = 2.23$; $\bar{X}II = 2.23$) and finally the English Control group ($\bar{X} = 1.60$) with the most errors. It is very encouraging to see that the two experimental groups scored favorably compared to the French group.

There was a significant difference between the IQ groups ($F = 18.25$; $df = 2, 38$; $p < .01$). The order of ratings from highest to lowest was high IQ group ($\bar{X} = 2.37$), average IQ group ($\bar{X} = 2.12$) and low IQ group ($\bar{X} = 2.05$).

Again, there was a significant difference between the sexes ($F = 32.05$; $df = 1, 17$; $p < .01$) with the girls ($\bar{X} = 2.31$) scoring higher than the boys ($\bar{X} = 2.05$).

There was no significant difference between the Markers' ratings.

Spelling 2 (Orthographe grammaticale): A summary of the mean ratings for each school and IQ group is presented in Table 16A. There was a significant school group effect ($F = 5.65$; $df = 3, 9$; $p < .01$). The French Control group was rated highest ($\bar{X} = 2.21$), followed by the two experimental groups ($\bar{X}I,II = 2.02$) and finally the English Control group ($\bar{X} = 1.50$).

TABLE 16A

SUMMARY OF RATINGS ON SPELLING II FOR FRENCH COMPOSITIONS

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	FRENCH CONTROL	TOTAL
High I.Q.	2.15	2.15	1.65	2.33	2.07
Average I.Q.	2.06	1.97	1.46	2.25	1.94
Low I.Q.	1.87	1.94	1.37	2.07	1.82
TOTAL	2.03	2.02	1.50	2.21	

TABLE 16B

SUMMARY OF RATINGS ON ORGANIZATION FOR FRENCH COMPOSITIONS

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	FRENCH CONTROL	TOTAL
High I.Q.	2.90	2.90	2.18	3.98	2.99
Average I.Q.	2.56	2.28	2.01	3.29	2.54
Low I.Q.	2.42	2.25	1.64	3.18	2.37
TOTAL	2.63	2.48	1.95	3.48	

TABLE 16C

SUMMARY OF RATINGS ON OVERALL FOR FRENCH COMPOSITIONS

	EXPERIMENTAL I	EXPERIMENTAL II	ENGLISH CONTROL	FRENCH CONTROL	TOTAL
High I.Q.	2.75	2.80	1.90	3.81	2.82
Average I.Q.	2.53	2.30	1.67	3.12	2.41
Low I.Q.	2.31	2.11	1.50	2.99	2.23
TOTAL	2.53	2.40	1.69	3.31	

There was also a significant difference between the Markers (F = 49.47; df = 1, 206; $p < .001$). Marker 1 ($\bar{X}$ = 2.09) tended to give higher ratings than did Marker 2 ($\bar{X}$ = 1.79).

There was no significant difference between the IQ groups nor was there a significant difference between the male and female students.

Organization: A summary of the mean ratings per school and IQ group is presented in Table 16B. There was a significant main effect due to School group (F = 40.44; df = 3, 17; $p < .01$). The French Control group ($\bar{X}$ = 3.48) was rated most highly, followed by the Experimental I group ($\bar{X}$ = 2.63), then the Experimental II group ($\bar{X}$ = 2.48) and finally the English Control group ($\bar{X}$ = 1.95).

There was a main effect due to IQ (F = 16.8; df 2, 22; $p < .01$). The high IQ group had the highest mean rating ($\bar{X}$ = 2.99), followed by the average IQ group ($\bar{X}$ = 2.54) and finally the low IQ group ($\bar{X}$ = 2.37).

There was a significant difference between the Markers (F = 41.99; df = 1, 206; $p < .001$). Marker 1 gave lower ratings ($\bar{X}$ = 2.47) than did Marker 2 ($\bar{X}$ = 2.79).

There was a significant Marker by Sex of student interaction (F = 6.15: df = 1, 206; $p < .05$). Marker 1 gave higher mean ratings to the female ($\bar{X}$ = 2.71) than to the male students ($\bar{X}$ = 2.28) while Marker 2 did not differentiate between them ($\bar{X}_{males}$ = 2.67; $\bar{X}_{females}$ = 2.88).

Overall: A summary of the mean ratings per school and IQ group is presented in Table 16C. There was a significant difference between the groups (F = 35.93; df = 3, 10; $p < .01$). The French group ($\bar{X}$ = 3.31) had the highest mean rating, followed by the two Experimental groups ($\bar{X}$ I = 2.53; $\bar{X}$ II = 2.40) and finally the English Control group ($\bar{X}$ = 1.69). These ratings give the French group a "C" standing, the experimental groups a "D" standing and the English Control group an "E" standing.

There was a significant difference between the IQ groups (F = 13.11; df = 2, 11; $p < .01$). The high IQ group had the highest mean rating ($\bar{X}$ = 2.82).

followed by the average IQ group ($\bar{X}$ = 2.41) and finally the low IQ group ($\bar{X}$ = 2.23).

There was a significant sex difference (F = 15.30; df = 1, 8; p <.01) with the girls ($\bar{X}$ = 2.66) scoring, on the average, higher than the boys ($\bar{X}$ = 2.31).

There was no significant difference in the way the Markers scored the papers for Overall (F = .44; df = 1, 206; p >.05).

Both raters agreed that among the experimental students the major problem was the incorrect use of verbs and verb tenses. The second most serious problem was gender. There was little agreement between the raters on the seriousness of the other items listed in question 18.

When we inspect the same data from the francophone students we see that again gender was regarded as the second most serous problem, but "incorrect use of verbs and verb tenses" was thought to be the third major problem area for the francophones. Thus, it appears that the most important grammatical problems for both the francophones and the immersion students tend to be the same. These findings must be accepted cautiously because there was a fairly high degree of inconsistency between the two French markers.

In summary, we find that on all six measures the French group performed best, the English Control group performed the poorest and the two experimental groups were intermediate - better than the English Controls but not as good as the French Controls. In order to understand what the ratings actually indicate it is necessary to interpret them with reference to the rating scales and criteria presented in Appendix C. Considering the lack of emphasis put on French writing skills in the immersion program the experimental students' performance on this test compares favourably with that of the French Control students.

French Oral Production Skills

In the past the children's oral production and listening comprehension skills have been rated subjectively on a number of different dimensions; for example, rhythm and intonation, punctuation, grammar, vocabulary, etc. In general, their ratings have indicated that the French-immersion students speak better than the English Control children but not as well as the French Control children. Consistent with these systematic reports, this year's testers reported anecdotally that the French-immersion students were more relaxed and willing to communicate in French than the Control students. The French-immersion students were quite adept at improvising in order to communicate their ideas. It was felt that another analysis of thissort would be repetitious and would not add any new, useful information. Rather it was felt that a more detailed linguistic analysis of the children's speech would be a useful supplement to the subjective ratings that have been made in the past.

Thus, some of the speech samples from Experimental I and English Control students' speech were analyzed by a linguist. Owing to the time required to do such analyses only ten students from the Experimental 1 group and the English Control group were examined. They were chosen randomly from the entire sample. Two types of analyses were performed on each student, a grammatical analysis and a phonetic analysis. Table 17 presents a list of the phonological items that were selected for evaluation and Table 18 presents a list of the grammatical structures that were analyzed. These items were selected because they were felt to be either important for effective communication or likely sources of difficulty for anglophones.

Method of Scoring:

To score each item, grammatical or phonological, the linguist rated the first three occurrences of that item as either correct or incorrect. If the item did not occur in the student's speech then no rating of it could be made. Thus, the scores represent the percentage of total items spoken that were spoken incorrectly. A summary of the phonetic analysis is presented in Table 17 and a summary of the grammatical analysis is presented in Table 18. It is evident that some of these percentage scores are based on small total numbers and a small sample of students. Interpretation of the findings must bear this in mind.

Phonetic Analysis:

The Experimental I Students made a smaller percentage of errors on only two of the items, Stress and Reduction. The English Control students tended to make smaller percentages of errors on the remaining items. These findings were contrary to expectation. It had been expected that because the experimental students had used French more than the Control students their phonetic skills would be better. One possible explanation of these findings is as follows: because of their French-immersion experience, the experimental students are concerned primarily with the communicative aspects of French, perhaps even at the expense of strictly proper French. The Control students, on the other hand, are less concerned with communicating in French (this is obvious from cursory observations) and are more concerned with the correct structural aspects of French. As well, constantly listening to non-native speakers of French, as the immersion students do, may result in poor phonetic skills.

TABLE 17

PHONOLOGICAL ANALYSIS OF SPEECH SAMPLES

PHONOLOGICAL FEATURE	EXPERIMENTAL I		ENGLISH CONTROL	
	TOTAL FREQUENCY OF USAGE*	PERCENTAGE ERROR	TOTAL FREQUENCY OF USAGE*	PERCENTAGE ERROR
1. /r/ Incorrect: English [ɹ] instead of [r̃], [ʀ]	30	40%	30	23%
2. /l/ Incorrect: English [l] Count only word-final [l]	30	50%	27	30%
3. Aspiration. Incorrect: Strong aspiration of p, t, k. (Count only in word initial p, t, k)	30	60%	30	50%
4. Stress: Stress at Proper Position (Count only words with 2 or more syllables that do not have a close cognate in English with a final stressed syllable)	30	37%	30	66%
5. Reduction. Unstressed vowels to [ə] or dropped, e.g. lapin [lapɛ̃] (count words with 2 or more syllables)	30	10%	29	24%
6. /y/ Incorrect: if /y/ is pronounced [jy] or [ø] as in musique [mjysik] or vu [vø] /ø/ Incorrect: if /ø/ is pronounced [œ] as in deux [dœ]	30	37%	30	20%
7. Diphthongization of o and e in word final position. e.g. beau [boʊ] (count only word final position)	30	20%	30	17%
8. Schwa Deletion. Retention of a schwa when it normally would be dropped, e.g. à demain [adəmɛ̃] (count only those schwa that normally could have been dropped in the expression you are listening to	30	60%	30	53%

* Maximum total frequency = 30

Grammatical Analysis:

On three of the items, Gender, Negation and Contraction the Experimental I students made a smaller percentage of errors than did the Control students. On two of the items, Personal Prounouns and Expressions of Time, the Experimental Students made a larger percentage of errors. However, this can be explained by the fact that the Control students did not use these constructions at all. The Experimental students were using more complex, grammatical constructions than the Controls and were thus prone to make more errors. On two of the items, Avoir/Etre and Reflexive Verbs, the Controls made smaller percentages of errors. On Tense Use the percentage error was approximately equal for both groups.

In summary, the Experimental students seem to be poorer than the English Control Students in the phonetic aspects of their French. This finding may be due to the Experimental group's concern for the communicative aspects of French rather than the structural, formal aspects of it. The English Control students, on the other hand, seem to approach French in a much more academic and formal way, emphasizing correctness instead of communication. On grammar the Experimental students seemed to make more errors than the Controls on complex grammatical constructions that the Controls avoided using, such as expressions of time and use of personal pronouns.

Discussion and Summary

On all tests of English-language skills the two experimental groups of students performed at a level comparable to the English Control group of students. In particular, there were no differences between the groups on standardized tests of spelling, vocabulary and reading. On the vocabulary test, both experimental groups scored favourably as compared to a larger American sample; namely, 70th percentile. On the reading test, their percentile standing was less favourable,

TABLE 18

GRAMMATICAL ANALYSES OF SPEECH SAMPLES**

ITEM	EXPERIMENTAL I		ENGLISH CONTROL	
	TOTAL FREQUENCY OF USAGE*	PERCENTAGE ERROR	TOTAL FREQUENCY OF USAGE*	PERCENTAGE ERROR
1. Gender Errors	30	57%	21	62%
2. Misplacement of personal prounouns	2	100%	0	0%
3. Proper Contraction (a les portes = aux portes)	26	38%	9	66%
4. Improper use of negation (pas, que, de after negation)	11	0%	9	11%
5. Tense Usage. The tendency to use the infinitive all the time (Je manger)	30	17%	27	15%
6. Confusion between avoir/être (in passe compose or idomatic)	21	24%	16	13%
7. Expressions of time, correct tense (il y a, depuis, etc.)	4	75%	0	0%
8. Proper use of reflexive verbs (se lever)	11	73%	6	17%

*maximum total frequency - 30 ** analysis is based on 10 students in each group

with the Experimental I students scoring at the 66th percentile and the Experimental II students scoring at the 50th percentile.

On all three measures, spelling, vocabulary and reading, the higher I.Q. groups scored higher than the low I.Q. groups. This is to be expected since I.Q. scores correlate highly with measures of verbal skills.

On writing, a new test introduced this year, the experimental students scored as well as the control students. All students, experimental and control alike, were rated "satisfactory", on measures of spelling, sentence accuracy, sentence complexity and variety, organization, originality and overall.

Again, there was a general tendency for the high I.Q. students to score higher than the low I.Q. students.

On the French tests, in general, the experimental students scored higher than the English control students but not as well as the Francophone controls. Specifically, the experimental students performed better than the English controls on the reading test, the Test de Rendement en Français, and the Test de Rendement en Mathématiques. On the Test de Rendement en Français the experimental students performed at a level comparable to the top stanine of the below average portion of the Grade V Francophone students in the C.E.C.M. On the Test de Rendement en Mathématiques, the experimental students performed at a level comparable to an average grade VII francophone group of students in the C.E.C.M. In fact, the experimental students scored as well as the French controls on this test. Bear in mind, however, that the norms for the C.E.C.M. were established in October, 1973, and, thus, the scores from this testing which was in February, 1974 reflect this additional four months.

Two different types of analyses of French oral production skills were tried this year. Detailed phonetic and linguistic analyses were made of speech samples of ten experimental students and ten English control students. These findings should be regarded as tentative owing to the smallness of the sample and to the novelty of the analysis. Nevertheless, analyses of this type might be instructive in pinpointing the students' linguistic problems so that they might be improved.

The findings from the phonetic analysis suggest, contrary to expectation, that the experimental students' phonetic skills are no better than those of the English control students. It was suggested that this might be due to the different concerns expressed by these two groups of students towards French. The experimental students may be more concerned with the communicative aspects of French as reflected by the ease and willingness with which they speak it. The English control students, on the other hand, may be more concerned with the formal structural aspects of French and with speaking French correctly than are the immersion students.

The findings from the grammatical analysis suggest that the experimental students are using more complex grammatical constructions than the controls, possibly because the experimental students are more communicative in French and are, therefore, trying to say more.

On writing, the experimental students were consistently rated higher than the English control students, on measures of vocabulary, spelling, sentence accuracy, organization and overall. The raters felt that the two most serious errors made by the experimental students were "incorrect use of verb tense" and "gender".

In summary, these findings are consistent with those from the two previous evaluations on the Grade VII group. The experimental students do not

appear to be experiencing any difficulties in their English-language skills. In fact, in some instances they were superior to the English control students despite the fact that IQ had been equated. In the area of French-language skills, the findings from the phonetic and grammatical analyses of spoken French suggest that there are well-defined problem areas that could possibly be systematically improved. In terms of communication skills, however, the experimental students are decidedly better than the English controls - they score better on reading and writing tests and, most obvious of all, they are more relaxed and willing to speak French. The experimental students lack the inhibitions and self-consciousness that characterizes the average student.

The French language skills learned in the Grade VII French-immersion program provide a firm base for more advanced French-language learning studies and greater functional communicative competance in French.

APPENDIX A

ENGLISH COMPOSITION

Student code number _____

(A) SPELLING:

Perfect, no errors ---------- 5
Three errors or less ---------- 4
Four, 5 or 6 errors ---------- 3
Seven or more mistakes ---------- 2
Unacceptable for a grade 7 student ---------- 1

PLEASE CIRCLE ERRORS

(B) PUNCTUATION:

Appropriate _____

Inappropriate _____

(C) LENGTH:

Appropriate: fully developed his ideas in appropriate length _____

Inappropriate: Far too long and rambling _____
or too short without completing ideas _____

(D) SENTENCE ACCURACY: grammatical correctness,
writing in complete, correct sentences

Excellent: consistent use of complete, well-formed sentences; no major errors ---------- 5

Very Good: some major errors; generally above standard -------- 4

Satisfactory: adequately meets objectives of grade 7 level ---- 3

Below Standard: many major errors; needs work ---------- 2

Unacceptable for grade 7 level ---------- 1

*UNDERLINE GRAMMATICAL ERRORS

(E) SENTENCE COMPLEXITY AND VARIETY:

Excellent: superior use of complex and varied sentence types in a meaningful way ---------- 5

Better than average use of complex and varied sentence types --- 4

Complexity and variety of sentence types is adequate for grade 7 level ---------- 3

Overuse of simple sentences: needs work ---------- 2

Far below grade 7 level ---------- 1

cont'd...

ENGLISH COMPOSITION page 2

(F) ORGANIZATION: ideas flow smoothly from opening sentence; ideas well-developed; use of paragraphing where appropriate; story conclusion.

Excellent: outstanding organization ---------- 5

Above average: coherent flow of ideas with good ending -------- 4

Meets Grade 7 standards: generally coherent flow of ideas with suitable enging ---------- 3

Below standard: at times incoherent; no ending ---------- 2

Totally inadequate: ---------- 1

(G) VOCABULARY:

Superior ---------- 3

Average for expectations of Grade 7 ---------- 2

Below average expectations ---------- 1

(H) ORIGINALITY:

Very imaginative ---------- 4

Above average ---------- 3

Ordinary ---------- 2

Dull and uninteresting ---------- 1

(I) OVERALL RATING:

A - Excellent

B - Above average

C - Average

D - Below average

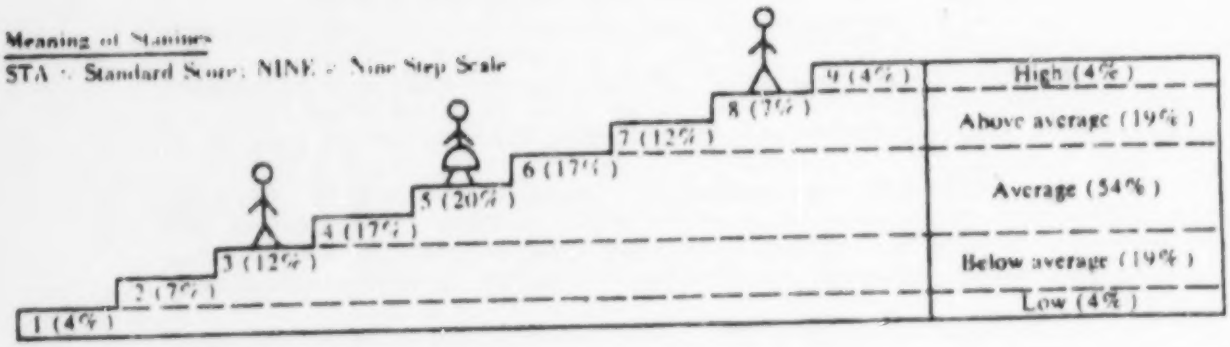
E - Unsatisfactory

APPENDIX B

INTERPRETATION OF STANINE SCORES

Stanine - The stanine scale is a simple nine-point scale of standard scores. (The word stanine was derived from STAandard NINE-Point scale). In this scale, raw scores are converted to scores which range from 1 (low) to 9 (high) with a mean of 5 and a standard deviation of 2.

FIGURE 1. Percentage of Cases at Each Stanine Level.



APPENDIX C

RATING SCALE FOR THE ASSESSMENT OF FRENCH COMPOSITION - Grade 7, Jan.1973-74

1.A. SENTENCE ACCURACY: Ability to write simple correct sentences; correct grammar, syntax, etc.

Excellent - native-like use of complete well-formed sentences. no major errors ---------- 5
Very good - some major errors, generally above standard ---------- 4
Satisfactory - adequately meets minimal requirements of written French ------ 3
Below standard for written French - needs work ---------- 2
Unacceptable written French ---------- 1

B. *Please rank the following grammatical faults in order of importance from 1 (most serious) to 5 (least serious):

- verb: correct use of tenses (excluding avoir & être)
- verb: continuity of appropriate tenses
- avoir/être: misuse - (e.g. j'ai alle; Je suis douze ans)
- avoir/être: correctly used
- syntax: correct word order (e.g. placement of personal pronouns, adjectives, etc.)
- gender of nouns
- pronouns: (correct form of pronoun; use of all personal pronouns including y/en)

2. VOCABULARY

Native like use and richness of vocabulary (e.g. ability to use synonyms)---- 5
Very good vocabulary for a non-native speaker. Generally uses the right words where appropriate ---------- 4
Sometimes uses inappropriate terms or anglisisms due to lack of adequate vocabulary ---------- 3
Frequently uses inappropriate words or anglisisms to express himself. Needs work ---------- 2
Vocabulary in French insufficient to express simple thoughts ---------- 1

3a. SPELLING (Orthographe d'usage) (Invariant word spellings)

Few errors, mostly minor ---------- 3
Acceptable - majority of words spelled correctly - between 10/15 errors ----- 2
Unacceptable, over 15 errors, or majority of words misspelled ---------- 1

3b. SPELLING (Orthographe grammaticale) Verb endings and agreements)

Few errors, mostly minor---------- 3
1 Acceptable - some major errors ---------- 2
Unacceptable - majority of words misspelled ---------- 1

4. ORGANIZATION OF MATERIAL/CONTINUITY OF IDEAS

Ideas flow smoothly from opening sentence, ideas properly developed with appropriate conclusion.

Excellent - outstanding organization ---------- 5
Above average - coherent flow of ideas with good ending ---------- 4
Generally coherent flow of ideas with suitable ending ---------- 3
Below standard - often incoherent, no ending ---------- 2
Totally incoherent ---------- 1

5. OVERALL ASSESSMENT

Excellent ---------- 5
Very Good ---------- 4
Good ---------- 3
Poor ---------- 2
Unsatisfactory ---------- 1

Additional comments if relevent - please be specific.

FOOTNOTES

1. Newman-Keuls Multiple-Comparison Test:

Newman-Keuls comparisons (or *q*) is a statistical procedure which compares two mean scores at a time and gives the probability level associated with the difference between those two means. This procedure contrasts with an *F* test or analyses of variance which usually compares more than two means at a time to one another and indicates the probability level associated with the combined differences between these means. For example, if you have mean scores for three groups - Experimental I, Experimental II and Control - a significant *F* test would indicate that there is a statistically significant difference between at least two of the three means, it would not indicate where the difference actually is - between Experimental I and II, for example, or between Experimental I and Control, etc. You would use a Newman-Keuls test, after the *F* test, to pinpoint which specific groups differ significantly from one another. Thus, you would do a Newman-Keuls or *q* test comparing each of the three means, one at a time, to each of the other means.

2. I would like to express my thanks to Dr. Y. Morin of the Department of Linguistics, University of Montreal, for his generous assistance in designing the analyses of French oral skills.